

RESEARCH ARTICLE

eLAMI—An Innovative Simulated Dataset of Electrical Loads for Advanced Smart Energy Applications

LUCA TARI¹, (Graduate Student Member, IEEE),
GIUSEPPE BERRETTONI¹, (Graduate Student Member, IEEE),
CARMINE BOURELLY¹, (Graduate Student Member, IEEE),
GIANNI CERRO², (Member, IEEE), DOMENICO CAPRIGLIONE¹, (Senior Member, IEEE),
AND LUIGI FERRIGNO¹, (Senior Member, IEEE)

¹Department of Electrical and Information Engineering, University of Cassino and Southern Lazio, 03043 Cassino, Italy

²Department of Medicine and Health Sciences, University of Molise, 86100 Campobasso, Italy

Corresponding author: Luca Tari (luca.tari@unicas.it)

ABSTRACT Smart Energy Applications are particularly impacting, especially due to energy resource scarcity and its high associated costs. Smart management of energy consumption derives both from the user lifestyle, in terms of efficient and responsible behaviors, and from automatic algorithms that control and counteract energy waste and inefficient management. Focusing the attention on the latter, the development of methodologies and well-working techniques to monitor and optimize consumption often requires an important effort in long measurement campaigns to get raw data to work with. Whenever this should be too much expensive or proper instrumentation is unavailable, public datasets could solve the problem. The current literature review on the dataset availability showed a large presence of information, especially related to electrical energy consumption. Nevertheless, several limitations affect them, from the low number of calculated electrical parameters (i.e. 4-5 in most cases) to short analysis periods, passing by the lack of detailed frequency domain information or poor consumption habit transitions analysis. Accordingly, this work aims to overcome current dataset limitations, by proposing a real-measurement based simulated dataset, extracting more than 400 discriminative electrical parameters on 36 different home appliances, discussing preliminary acquisition set-ups, simulation process, extracted electrical parameters and examples of applicability to smart energy applications. To provide a data quality index, a validation procedure has also been carried out, showing how simulated data match real acquisition with a reference measurement instrument. The produced dataset is available for downloading and analysis in public free access and its repository link is provided in the reference section.

INDEX TERMS Smart energy, dataset, energy consumption simulator, smart home, load profiling, non-intrusive load monitoring, energy management systems.

I. INTRODUCTION

Applications related to the smart energy paradigm are demanding novel devices and techniques for satisfying the increasing issues of more efficient and sustainable use of

The associate editor coordinating the review of this manuscript and approving it for publication was Neetesh Saxena¹.

energy. Nowadays, these applications are attracting more and more interest, especially due to energy and ecological transitions, whose goals are the reduction of CO₂ emissions and the increase of energy efficiency. As the latter increases, international treaties set limits on emissions and global temperature rise that must be met, such as those defined at COP26 [1]. As an example, in Europe, the development and use of

intelligent measurement systems and techniques has been boosted by European directives in recent years. In 2006, the Directive 2006/32/EC [2] identified the use of intelligent metering systems as one of the tools to improve energy efficiency through suitable algorithms and techniques. In 2009, the Electricity Directive 2009/72/EC [3] declared the obligation for the Member States to ensure the adoption of smart metering systems. Finally, in 2012 the European Energy Efficiency Directive 2012/27/EU [4] reaffirmed the importance of using smart meters, with strict requirements for the Member States regarding both metering and billing. In this new paradigm, in which energy efficiency and the development of new services based on the measurement of energy parameters are demanding data, the smart meter assumes a key role. On one hand, it accomplishes the primary task of monitoring energy consumption. On the other hand, it could provide new quantities useful for modern smart energy applications. Among these quantities we can mention those related to the electrical signature of a device, the power quality, and to the use of energy trends over time, to cite a few, which could enable the implementation of new high-value services for energy and plants management. In other words, the smart meter can be the key-enabling technology to implement modern algorithms and emerging techniques in the field of smart energy applications such as loads defragmentation through the non-intrusive load monitoring (i.e. NILM), digital twin of the equipment, clustering of devices and predictive detection and diagnosis of faults on both plants and grids. More in-depth, the above-mentioned techniques are generally based on machine learning algorithms or on optimization techniques exploiting electrical parameters related to the electrical signature of a load and/or Power Quality parameters. For example, considering the development of NILM solutions that is involving many researchers in the scientific community, some algorithms are based on the use of Markovian models (HMM) and their variants [5], [6] [7], while others prefer signal processing techniques using graphs (Graph Signal Processing) [8], [9] or Combinatorial Optimization [10]. In recent years, other Machine Learning techniques have also been applied for non-intrusive monitoring, such as Multilayer Perceptron (MLP) [11], Convolutional Neural Networks (CNN) [12], Deep Learning [13], [14], Recurrent Neural Network (RNN) [15], Extreme Learning Machine [16], and Bayes Classifier [17]. Whatever the followed approach, all these techniques require two fundamental steps to be implemented to reliably train and tune the algorithms on the considered case study: *a*) the accurate measurement of the energy consumption and parameters related to the power signatures and other parameters relying on the involved devices, *b*) the monitoring for a time interval wide enough for capturing an adequate amount of data. As for the first step *a*), it is crucial since it strictly affects the quality and the quantity of the collected information. It is very important to have energy measurements characterized by low uncertainties and, at the same time, to provide several energy parameters, such as the harmonic and inter-harmonic power,

current and voltage values, the total harmonic distortion, the power factor, etc. that can characterize how a device, or a group of devices, is consuming the electrical energy. As for *b*), the monitoring for wide time intervals is very important not only to provide big data to the smart energy algorithms but also to let the algorithms work on a few operating conditions able to give reliable information on the monitored process or equipment. For example, to optimize the energy consumption of a family it is important to monitor the consumption habits considering several days and the effects of seasonality. As a further example, such as diagnosis or predictive maintenance of devices and grids, to predict faults on equipment, it is important to have a reliable footprint of its energy status considering all the possible operating conditions of the involved device. Since the above-mentioned requirements impose to perform very expensive and time-consuming measurement campaigns, in the last years some datasets have been proposed in the literature with the aim of facilitating the development of new algorithms addressing the cited emerging needs. Many are the papers that propose the use of a dataset to train or tune smart energy algorithms. Examples for NILM can be found in [18], load profiling can be found in [19] and predictive maintenance can be found in [20]. In literature two different types of datasets can be found: real [21], [22], [23], [24], [25], [26], [27], [28], [29], [30], [31] and simulated datasets [32], [33], [34], [35]. In particular, real datasets are composed of data, collected on the field, while simulated ones are obtained through suitable software simulators. However, these datasets do not meet all the requirements for effective data usage, i.e. the limited amount of saved electrical parameters, data variability, seasonality, closeness to real scenarios, and definition of the current operating states are just some of the missing or not complete information that makes them poorly suitable for modern Smart Energy Applications. To overcome such limitations, in this paper we present an innovative energy dataset. The main advantages of the proposal are: *(i)* data variability, in terms of the operating states of the electrical loads and the adoption of appropriate consumption models; *(ii)* the total number of available electrical parameters (433 in our case) enormously larger than the above-cited datasets, as the full analysis of frequency behaviors and harmonic computations; *(iii)* the seasonality of the data. Furthermore, the presented dataset is a hybrid solution between the simulated datasets and the ones based on real data. Indeed, it is generated through simulation, starting by real measurements, and customized by means of an effective computed variability model. Finally, it has been validated with further *a-posteriori* measurement campaign on real loads. The developed dataset, namely eLAMI, is made publicly available to the whole research community at [36] ([download here](#)) through a modular structure, allowing customized downloads and analyses.

The structure of the paper is the following: related works about some different public datasets is reported in Section II; Section III describes the design of dataset generations. The results of the proposed dataset are reported in

TABLE 1. Main characteristics of some public dataset based on real data.

Name	Number Appliances	Parameters	Period
AMPds [21]	20	P,Q,S,V,I	1 year
AMPds2 [22]	20	P,Q,S,V,I	2 years
BLUED [23]	50	P,Q,V,I	1 week
Dataport [24]	8598	P,S	4 years
DRED [25]	12	P	6 months
ECO [26]	45	P,V,I	8 months
ENERTALK [27]	75	P,Q	1714 days
iAWE [28]	63	P,Q,S,V,I	73 days
REDD [29]	92	P,V,I	119 days
REFIT [30]	20	P	2 years
UK-DALE [31]	109	P,Q,S,V,I	2247 days

Section IV and finally in Section V, final considerations are provided.

II. RELATED WORKS

The growing demand of new monitoring devices with smart algorithm and technique has led to the development of several dataset capable to tune the performance. In Tab. 1, some examples of public datasets with data collected from the field are shown; they differ for the number of devices, number and type of provided electrical parameters and length. In the following they are briefly described.

- **AMPds** [21] The Almanac of Minutely Power dataset is a public dataset, published in 2013, containing 1 year of collected data of residential appliances from a single household in Canada. This first version contains measurements of electricity, water and natural gas at one-minute intervals, for a total of 525600 readings per year per meter.
- **AMPds2** [22] This second version of the AMPds dataset differs from the previous one only in the number of total readings, 1051200, corresponding to 2 years of acquisition.
- **BLUED** [23] The Building-level fully labeled dataset for electricity disaggregation was released in 2012 by 1 week electricity data from 1 building in the USA. This dataset contains not only the steady-state, but also the state transition of each appliance.
- **Dataport** [24] The Dataport database was created by Pecan Street Inc and published in 2015. It contains electricity data from 722 houses and commercial buildings across different cities in the USA. As it has a sampling period of 1 min for aggregate and appliance signal, this is considered a low frequency dataset.
- **DRED** [25] The Dutch Residential Energy Dataset was released in 2015 and contains energy consumption data from a household in Netherlands, with a total duration over six months. It includes electricity measurements for the aggregate and submetered signal of each device.
- **ECO** [26] Electricity Consumption and Occupancy dataset was collected in 6 Swiss households over a period of 8 months. It contains data recordings of active power, voltage and current at low frequency sampling rate.

- **ENERTALK** [27] The ENERTALK dataset was created in Korea from 22 houses with a total period of 1714 days and it was published in 2019. It provides active and reactive power measurements (both aggregate and each device), with a sampling frequency of 15 Hz.
- **iAWE** [28] Indian Dataset for Ambient Water and Energy was released in 2013, from recordings electricity, water and ambient data in a house in New Delhi, for a total duration of 73 days. The electrical data were recorded with a sampling period from 1 to 6 seconds over 63 electrical appliances.
- **REDD** [29] Reference Energy Disaggregation Dataset has been published in 2011. It contains 119 days of collected data from 6 households in the USA and includes both high and low frequency recordings.
- **REFIT** [30] The REFIT Electrical Load Measurements dataset includes cleaned electrical consumption data from 20 households in the UK from 2 years of recordings in 2016. It contains electrical data with a sampling period of 8 s and active power as only parameter.
- **UK-DALE** [31] UK-Domestic Appliance Level Electricity was published in 2015 and it contains 2247 days of data by 5 residential buildings in the UK. Just like REDD, it reports high and low-frequency data and all appliances are sub-metered.

Collecting data to build a dataset is a fairly complex process. As long measurement campaigns have to be carried out, the process requires a considerable amount of time, effort, and instrumentation to measure and record data. One of the main limitations of the real datasets in the literature today is the small number of reported electrical parameters (at most P, Q, S, V, I).

This could represent a limitation for algorithms in the field of Smart Energy. In particular, considering methods for load profiling, NILM and fault or predictive diagnosis, the optimal choice of needed parameters is still a research topic: therefore the availability of a large number of electrical quantities, able to define the complete “electrical signature of the load”, could help in performing an accurate selection.

Furthermore, real datasets are often characterized by missing data due to several problems that may occur during the measurement campaigns, time mismatching between individual load data and aggregate data or possible errors due to malfunctioning and inaccuracy of the adopted instrumentation.

The absence of information about current operating states of each device can also be a limitation for these datasets, e.g. in the case of supervised artificial intelligence algorithm training or in the definition of consumption quality as well as system efficiency indices. A solution to the aforementioned problems could be the simulation of data.

A simulated dataset does not require lengthy monitoring campaigns, saving time, costs, and instrumentation. All these advantages are provided as long as the simulation process is correctly implemented, which is not trivial in terms of suitable modeling and computational costs.

TABLE 2. Main characteristics of some public dataset based on simulated data.

Name	Number Appliances	Parameters	Period
AMBAL [32]	14	-	1 days
SHED [33]	66	P, S, V, I	14 days
SmartSim [35]	25	P	7 days
SynD [34]	21	P, S, V, I	180 days

In literature, some simulated datasets are presented as reported in Tab. 2 and discussed below.

- **AMBAL** [32] Automated model builder for appliance load dataset was published in 2017 and it comprises 14 domestic loads at a sampling rate of 1 Hz for a time duration of one day. The AMBAL dataset allows the user to build models using real energy consumption data, based on parameterized signature sequences. The main operational phase of the AMBAL dataset includes pre-processing, extraction of active segments, segmentation, and model fitting.
- **SHED** [33] Simulated high-frequency energy disaggregation dataset, released in 2018. It is a commercial dataset containing the power consumption of 66 buildings at a sampling frequency of 1/30 Hz. The data is generated synthetically and based on modeling the current flowing through an electrical device and is matched with the real model of electrical devices.
- **SynD** [34] SynD is a synthetic dataset that was published in 2020 and simulates readings of electricity consumption for a house for 180 days. The measurements campaign was based on the monitoring of 21 different residential devices from 2 households in Austria. In particular, the consumption patterns were observed, the absorption profiles of each device were then extracted and finally the dataset was generated.
- **SmartSim** [35] A Device Accurate Smart Home Simulator for Energy Analytics was released in 2016 and is a simulated dataset of 1 week total duration. It uses the energy modelling of individual devices to build the final dataset with the aim of generating accurate domestic energy traces that are qualitatively and quantitatively similar to real energy data traces.

Despite the large amount of simulated datasets in the literature, most of them still suffer some of the aforementioned problems, as limited number of saved electrical parameters, absence of harmonic and power quality information of seasonality, monthly, daily variability, as well as variability of operating states; furthermore, they often provide a low likelihood value with respect to real scenario profiles.

III. DESIGN OF DATASET GENERATION

In this section, the design process of the eLAMI (electrical Loads Acquisition for Monitoring Instruments) dataset and its implementation is described. At first, main requirements that a modern energy dataset must have are discussed. Issues related to the the choice of a simulator and possible solution

are then faced. Chosen electrical loads, along with consumption patterns and the simulator description close the section.

A. REQUIREMENTS

A modern energy dataset for monitoring and investigating consumers' energy behaviors must have specific characteristics. The main requirements are:

- High number of saved electrical parameters, both in time and frequency domain;
- High likelihood to real scenario profiles;
- Faithful representation of the devices;
- Considering the metrological performance of commonly used energy smart meters;
- Presence of Power Quality and harmonic data;
- Appropriate observation times congruent with the objectives;
- Current operating state of the monitored device.

As regards *a)* having a large number of electrical parameters allows for a better representation of the electrical signature of the load. In this way, for example, as also proved in Section IV, energy efficiency algorithms can drastically increase their performance. Therefore, in the eLAMI dataset, 433 electrical parameters are calculated at each measurement interval. Furthermore, in our simulated dataset we try to represent an electrical scenario as faithful as possible to reality. Specifically, reported consumptions are related to a 2-person-household.

Having a good match between the simulated scenario and the real one, simply summarized as likelihood, is crucial as it offers the possibility, for instance, of training AI algorithms on consistent data, as mentioned in *b)*. This avoids possible mismatches between performance obtained in simulation and real scenarios.

As said in *c)* it is necessary to take into account both how the electrical energy is consumed, but also how the individual load behaves in terms of electrical operation in reality. In particular, in our case we have chosen to represent equipment as 'state machines', simulating the corresponding nominal operating states. We consider this choice to be valid, having chosen to simulate the assumed scenario under permanent regime. Of course, in reality, between different 'operating states' there are transients that can lead to more or less marked variations in electrical quantities. As regards *d)*, furthermore, these variations can also be related to the natural duty cycle of the equipment or be due to the uncertainty of the monitoring system used [10]. eLAMI reports this variability thanks to the mathematical model for generating absorption profiles implemented.

As said in *e)* today, given the widespread use of electronic equipment, we consider the study of the harmonic behavior of electrical loads to be of great interest. For example, the analysis of the frequency spectra of voltage and current absorption of devices can provide useful information on the health of loads and in general of the entire system, also in terms of the quality of the power supply system (Power Quality). eLAMI

in this case reports a considerable amount of harmonic and quality parameters.

With regard to the observation time of the monitored system f , it must be chosen in such a way as to meet the objectives of the applications for which it was hypothesized. Classification, clustering, NILM, Load Profiling, and energy retrofit algorithms in most cases aim at analyzing the system over sufficiently long time horizons. In our case, eLAMI refers to a time horizon of one year. As regards g , an operating state of the monitored device can be defined as a steady-state voltage and current joint profile, whose availability allows the dataset user to evaluate the proper load working cycle and extract electrical signature quality indices and detecting possible incoming anomalies.

Indeed, referring to “electrical signature”, it not only deals with the typical quantities (P , Q , I_{RMS} , etc.) but with the entire frequency spectrum of voltage and current profiles absorbed by the electrical load. Such information are therefore also state-related, i.e. they can change for each operating state of the load.

Therefore, the proposed simulator must be designed to generate voltage and current profiles under all possible load states tested and for a predefined simulated time, in order to get the spectrum information.

An issue raised at this stage is related to the way the frequency spectrum could be faithfully simulated. The approach here followed is to make measurements on real loads under different tested operating states and adopt acquired information as a basis to generate simulated profiles.

Based on the knowledge of these quantities and the known usage habits of the electrical loads, it is possible to create the dataset by using the acquired information.

Any accidental failures or malfunctions that might occur in the real electrical system are neglected in the current status of the simulator. This is reasonable as these are very rare events in reality that, when compared with the simulation time frame, can be neglected. The simulator modular structure would eventually permit to add such situations in a fairly easy fashion.

B. ELECTRICAL LOADS DESCRIPTION

According to the aforementioned requirements of a new simulated energy consumption dataset for the residential appliances with innovative saved electrical parameters, namely eLAMI, has been developed. The choice of simulating a residential building consumption profile has the aim to provide a means with innovative characteristics compared to the datasets currently present in the literature for the evaluation of new techniques and algorithms in the field of Smart Energy, including NILM, Load profiling, Management Systems, and Energy Efficiency algorithms.

eLAMI refers to 360 days of simulation of a house having 36 connected appliances, as described in Tab. 3. But in the future other years will be added. The “ID” is the appliance identifier, N_S is the overall discrete number of tested operating states (including the OFF state) and the P_{Nom} is the upper

TABLE 3. Electrical loads specifications. Returns the appliance identifier (ID), the name of the loads (Appliance), the number of states tested (N_S) and and the maximum active power that each load can absorb (P_{Nom}).

ID	Appliance	N_S	P_{Nom} [W]
1	Beard Trimmer	2	4
2	Blender	2	60
3	Boiler	2	220
4	Coffee Machine	2	1200
5	Dehumidifier	2	190
6	Desk Lamp	4	10
7	Electrical Oven	3	2400
8	Electric Toothbrush	2	5
9	Fan	4	40
10	Fan Heater	4	1900
11	Fridge	2	140
12	Hair Clipper	2	6
13	Hair Dryer	3	420
14	Hair Straightener	2	65
15	Headphones Charger	2	5
16	Hoover	3	600
17	Infrared Heater	2	1350
18	Iron	2	1300
19	Kettle	2	2000
20	Lamps Bathroom	2	30
21	Lamps Bedroom	2	90
22	Lamps Closet	2	75
23	Lamps Corridor	2	55
24	Lamps Entrance	2	65
25	Lamps Kitchen	2	125
26	Lamps Livingroom	2	145
27	Microwave	2	1300
28	Notebook	2	60
29	PC-Desktop	2	300
30	Printer	2	750
31	Router	1	5
32	Smart TV	2	105
33	Smartphone Charger	2	25
34	Toaster	2	640
35	TV	2	85
36	Washing Machine	4	800

bound each appliance can absorb, according to the manufacturer’s indications. At each measurement interval, 433 electrical parameters are computed, whose details are reported in Subsection III-D. In addition to the information on the individual load, the same parameters are also calculated for the total aggregate and three partial sub-aggregates consisting of subsets of loads. In addition to the calculated quantities, current operating state is also provided.

The partial aggregates are provided for the three different identified zones of the virtual house. For each zone, a subset of loads was defined. The combination of loads for each sub-group was chosen to obtain aggregates with a progressive number of loads, as shown in Fig. 1. This is an important feature for the structure of eLAMI as it offers researchers the possibility to test the algorithms on an increasing number of loads, therefore on an increasing complexity level.

Although all loads are connected in parallel, their supply is carried out by means of radial lines. Therefore, in eLAMI, the voltage at the terminals of each load is not the same but depends on the load conditions.

Moreover, the reported OFF state, for some devices, is an indication of a standby state. Therefore, the absorbed current is slightly different than zero.

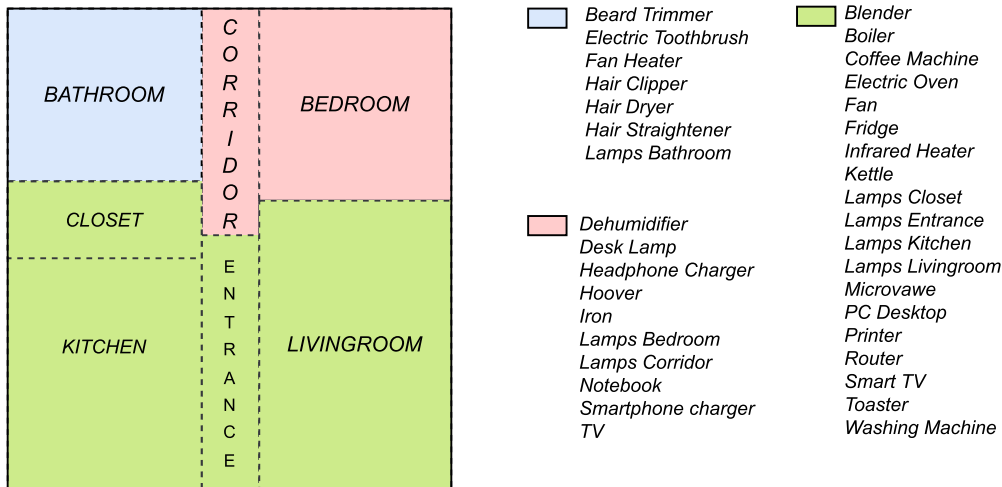


FIGURE 1. House Scheme and Electrical Loads Distribution. The blue zone represents the “Partial 1” with 7 electric loads; the pink zone represents the “Partial 2” with 10 electric loads; the green zone represents the “Partial 3” with 19 electric loads.

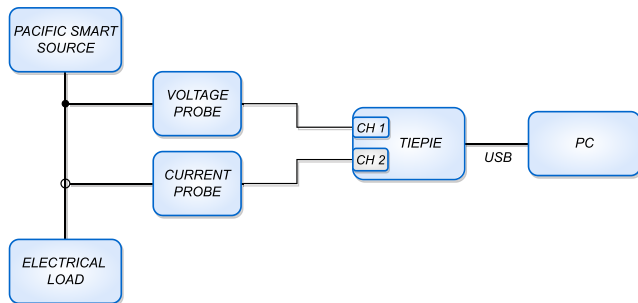


FIGURE 2. Block diagram of data acquisition setup.

C. SIMULATION BASIS: DATA ACQUISITION

In order to define the frequency spectrum of a load, as mentioned above, it is necessary to experimentally acquire its absorption profile. Through a measurement campaign in Industrial Measurement Laboratory (LAMI) in the University of Cassino and Southern Lazio, 36 residential typical electric loads (Tab. 3) have been acquired. Current and voltage waveforms for each devices has been recorded and get the corresponding reference profiles for use in data generation software. The block diagram of the experimental setup for data acquisition is shown in Fig. 2.

The adopted power supply system is a Pacific Smart Source, an electrical network emulator which allows reproducing any mains profile both in terms of amplitude and harmonic content [37]. In particular, it has been used as arbitrary voltage generator to supply the electric loads.

In order to emulate real working conditions, the power grid voltage was first acquired and the corresponding harmonic characteristics were calculated up to the 50th harmonic order. The harmonic coefficients obtained, in terms of amplitude and phase, are used as an input for the Pacific Power Source.

The emulated mains voltage profiles both with and without load, are shown in Figs. 3, 4 along with the harmonic

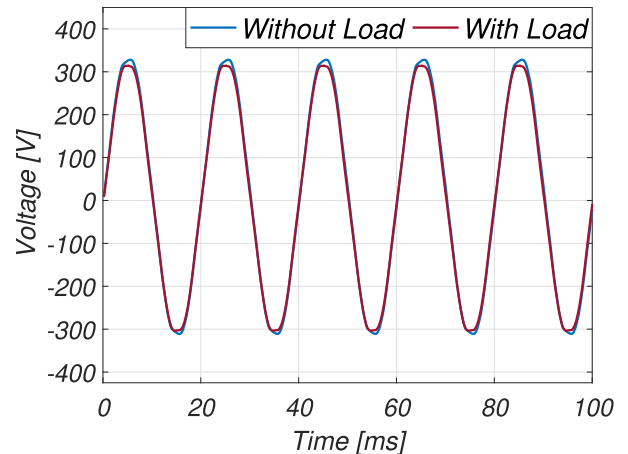


FIGURE 3. Emulated Voltage profiles with load and without load.

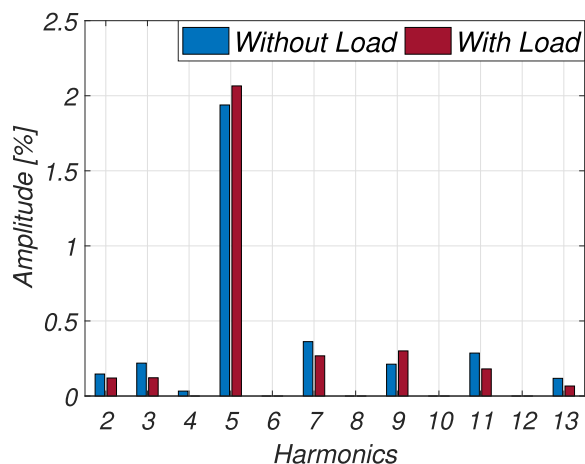
contents. For clarity we report the distribution of the percentages harmonic coefficients from the 2th to the 13th order. The harmonic coefficients shown in Fig. 4 were obtained by extrapolating the characteristics of the harmonics of the voltage signals and evaluating the amplitudes in percentage terms with respect to the corresponding fundamental tones.

Furthermore, it can be seen from the figures that there are differences between the voltages in different load conditions. In particular, the voltage under load is lower and has a slightly different distribution of harmonic coefficients. As expected, the voltage profiles of Fig. 3 do not reproduce a perfect sinewave, since it contains the harmonics contributions.

In Tab. 4, is reported a numerical comparison between the main acquired voltage and the emulated one, both with and without load. The compared values are: the root mean square (RMS) value (V_{RMS}) of both the overall voltage and only the voltage first harmonic ($V_{RMS \text{ Fundamental}}$). Furthermore, the offset component (V_{DC}), the peak value (V_{PK}), and the total harmonic distortion (THD) voltage are also compared.

TABLE 4. Comparison of some characteristics of the voltages: Acquired main voltage, emulated voltage with load and emulated voltage without load.

	V_{RMS} [V]	V_{RMS} Fundamental [V]	V_{DC} [V]	V_{pk} [V]	THD _v [%]
Acquired Main Voltage	231.35	230.82	12.79	326.27	3.94
Emulated Voltage With Load	232.27	231.99	10.06	327.31	4.85
Emulated Voltage Without Load	226.89	226.39	10.07	317.86	4.72

**FIGURE 4.** Percentage harmonic coefficients normalized with respect to the fundamental frequency.

Through the use of a Tektronix P202A Hall effect probe [38], powered by a Tektronix 1103 power supply [39], with a transformation ratio of 100 mV/A, the electrical current flowing in the circuit was measured.

Using a Tektronix P5200 differential probe [40], with a transformation ratio of 1:500, the voltage supplied by the Pacific to the electrical loads was measured.

A TiePie HS5 [41] was used to acquire the measurements through a customized software developed in MatlabTM environment. 30 repeated measurements (N_{ACQ}) were performed for each operating state tested. The iterated procedure allowed obtaining the corresponding average absorption profiles and the associated standard deviations for voltage and current waveforms.

The sampling frequency (F_s) used for profile acquisition is 5 kHz. For each acquisition (of each load state) 25000 points (N_P) were acquired.

The amplitude resolution of the acquisition system, through the TiePie Hs5, was set to 14 bits per channel, with a full scale of 0.8 for the voltage channel and 2 V for the current channel. Considering the conversion factors of the probes used, therefore (1:500 and 100 mV/A) this gives 400 V for the voltage channel and 20 A for the current channel. Values are chosen in relation to the electrical systems considered. All the sampling characteristics are summarized in Tab. 5.

D. SIMULATOR DESCRIPTION

The purpose of this section is to provide a general description of some of the fundamental parts that make up the innovative simulator created for the realisation of eLAMI, highlighting

TABLE 5. Sampling characteristics used for the acquisition of electrical load profiles.

Specifications	Descriptions
F_s	5 kHz
N_P	25000
T_{OSS}	5 s
N_{ACQ}	30
A/D converter resolution	14 bit
Full scale of the voltage channel	0.8 V
Full scale of the Current channel	2 V
Voltage probe conversion factor	1:500
Current probe conversion factor	100 mV/A

the procedures followed in the simulation. All processing operations were performed in MatlabTM environment.

The simulator consists of three sections: (1) INPUT, (2) PROCESSING and (3) OUTPUT.

For each section, the chronological flow and the main operations are described.

- 1) The “INPUT” section consists of 4 macro blocks:
 - 1.1) In “Simulation Parameter” block, specifications for the simulation are defined, e.g. simulation interval (1 year), measurement time (5 s), and aggregation criteria, just to cite a few. The content in round brackets is the assignment of possible parameter values. They are sent to the following blocks, defining the characteristics of the electrical scenario to be simulated.
 - 1.2) In “Electrical Loads Stochastic Model”, mathematical relations about operating state changes are defined according to output got by previous surveys performed by the authors in residential buildings.
 - 1.3) In “Definition of Consumption Habits”, taking into account the many factors influencing the electric consumption and adopting the mathematical models defined in the previous block, along with the hypothesized simulation interval, typical consumption habits, or patterns, are defined and used as input to generate faithful absorption profiles. The behavioural consumption patterns, defined for each load, take into account daily variability and seasonality. This is made possible by the implementation of a stochastic process, studied ad-hoc for the assumed dynamic system. To explain the suitability of the implemented stochastic model to avoid a mismatch with real scenarios, in our simulation framework it is absolutely unlikely that the blender or Hoover will switch on during the night as well as

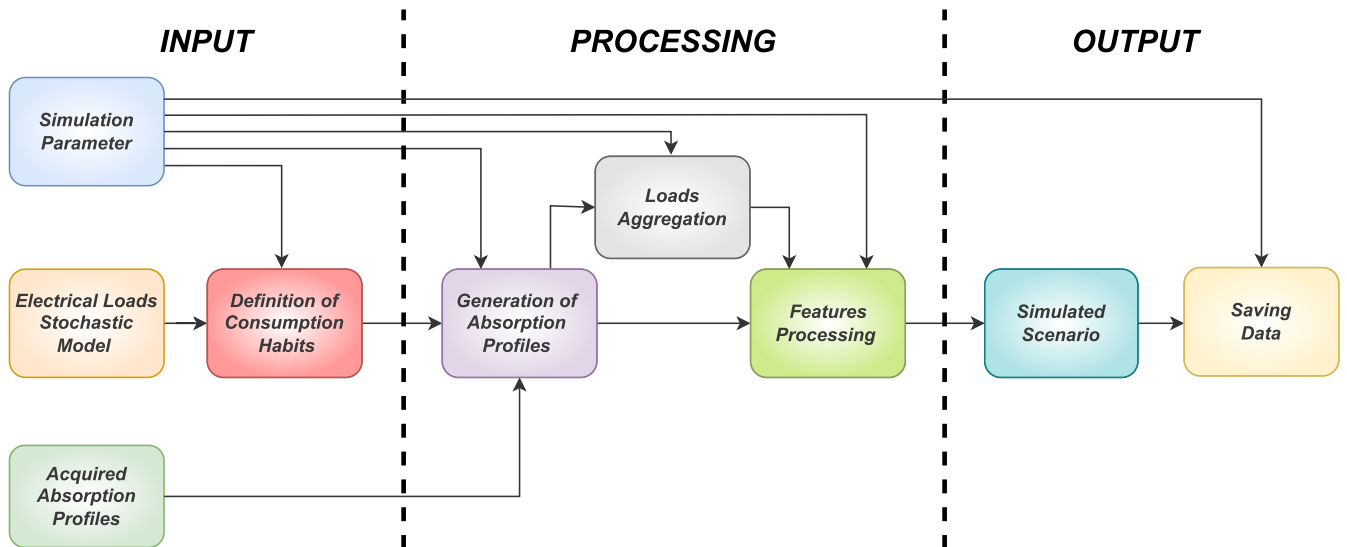


FIGURE 5. Block diagram of the simulator composed of three sections: “INPUT”, “PROCESSING”, and “OUTPUT”.

lighting or heating during winter periods for consecutive days is more likely than in summer times. In eLAMI, the defined stochastic model also considers all these factors. In this way, for each load, the days turn out to be dependent.

- 1.4) In “Acquired Absorption Profiles” contain the reference absorption profiles obtained as described in Section III-C.
- 2) The PROCESSING section is composed of:
 - 2.1) “Generation of Absorption Profiles” is a block that, taking inputs as described before, generates the voltage and current waveforms related to the conditions to be simulated. Such signals are simultaneously sent to both “Loads Aggregation” and “Features Processing” to perform different operations, as described below. Such waveforms are referred to the i -th iteration for a specific instant of a simulated day and condition of each considered appliance. The compounded values define the “Current Operating State” of each load.
 - 2.2) The “Loads Aggregation” block receives the output of “Generation of Absorption Profiles” and “Simulation Parameter” and aggregates individual loads in a macro-load condition, i.e. considers all appliances belonging to a specific category (e.g. bathroom) as they were one only aggregated load.
 - 2.3) In “Features Processing”, all electrical quantities assumed for individual loads and aggregates are calculated. For electrical parameters calculations, our simulator implements the definitions of Power in IEEE-1459, especially referring to single-phase non-sinusoidal case [42]. The calculated electrical parameters refer to: V_{RMS}^{TOT} and I_{RMS}^{TOT} , V_{RMS} and I_{RMS} at the fundamental, of harmonics and DC components alone; active,

apparent, non-active and distorted power; power factor and harmonics distorted parameters, for a total of 26 electrical parameters. Furthermore, for better identification of electrical loads signature, the simulator also implements the definitions of measurements in IEC 61000-4-7 Standard [43], regarding harmonics and interharmonics, including the rms values of voltage and current groups up to the 50th harmonic order are calculated, including harmonic sub-groups ($SubG_rV$, $G_rV_1 \dots G_rV_{50}$; $SubG_rI$, $G_rI_1 \dots G_rI_{50}$) and the corresponding phase values of each group calculated for a total of 202 parameters. In addition, the harmonic index of the maximum amplitude tones of voltage/current in each group is also provided, the phase of each harmonic group and the current operating state, to achieve further 205 parameters. Considering the overall processing, the total number of electrical parameters reported in eLAMI is 433.

- 3) The “OUTPUT” section is composed of:
 - 3.1) In “Simulated scenario”, the computed electrical parameters are packed considering the “simulation interval” (e.g. 1 day, 1 month, 1 year) parameter and their size also depends on the measurement time (e.g. 5 s), which is the time resolution over which the 433 parameters are computed. Future simulated years of eLAMI will be added to the main folder.
 - 3.2) In “Saving Data” block, a hierarchical structure is created for saving the dataset, as illustrated in Fig. 6. The latter was created to make the data as much usable as possible for the end user. From a hierarchical point of view, eLAMI is divided into a first level “by months”, then “by loads” and finally “by calculated electrical parameters”.

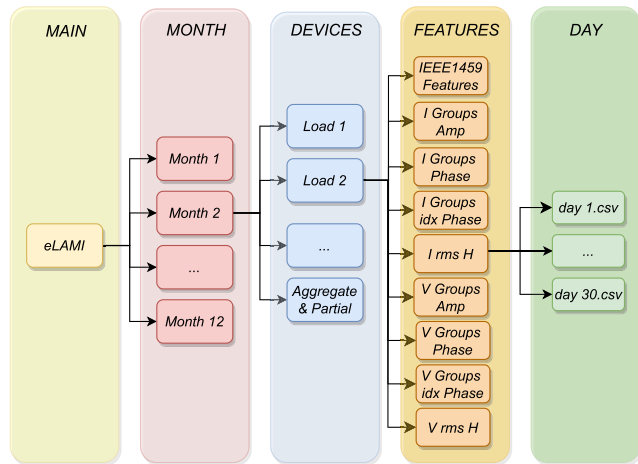


FIGURE 6. Saving eLAMI structure.

All information are stored in granular “.csv” files, one for each basic condition (day of the month).

IV. RESULTS

The aim of our work is to provide a dataset with innovative features compared to datasets currently found in the scientific literature, for the evaluation and development of new techniques and algorithms in the field of Smart Energy Applications. Of course, it is of paramount importance that the dataset is physically consistent with real case scenarios.

To this end, we first validate our dataset simulator; then, some dataset peculiarities are highlighted, and finally some examples of eLAMI applications in the field of Smart Energy, in particular Load Profiling, NILM, and Energy Management systems, are proposed.

A. VALIDATION OF DATASET SIMULATOR

For the technical validation of the simulator and the corresponding consistency of the generated data, a comparison between the measurements obtained from a real test and a simulated one has been carried out, by assessing their metrological compatibility. In particular, the electrical scenario assumed for the test, is composed of 3 real loads of eLAMI with different electrical characteristics, namely “Fan”, “Fan Heater” and “Smart TV”. They have been connected to the Pacific network emulator [37], then fed in parallel with the same voltage signal used for the creation of eLAMI reference profiles III-C. At the same time, the absorption profiles, at the ends of every single load and the “Aggregate”, were monitored using a laboratory wattmeter, the Precision Power Analyzer WT3000 [44]. The same scenario, without WT3000 measurement instrument, has been replayed in the simulation environment, by starting from the reference profiles previously acquired. Test set-up settings are total test duration (1 hr), measurement time (5 s), and the total number of measurement points (720).

At the end of the test, a comparison was made between the results obtained in the two cases, with the aim of showing:

i) a comparison between variability ranges in the case of real and simulated data; *ii*) the metrological compatibility of the simulated measurements, and thus of eLAMI, with real acquired values; the combination of *i*) and *ii*) leads to state the validity of the simulator implementation.

Starting from *(i)*, a comparison between the values obtained in the real and simulated cases is shown in Fig. 7, in terms of variability ranges. In this figure, the behaviour of V_{RMS} (7.a), I_{RMS} (7.b), P (7.c) and S (7.d) are reported, for each individual considered load and its corresponding aggregate.

In this case, the variability ranges are overlapped, although WT3000’s related range is almost always narrower than the simulated one. This is because the WT3000 has a higher accuracy level than what can be obtained by adopting the set-up used in the acquired reference profiles for simulation. In any case, since WT3000 has been chosen as a reference instrument, it is expected that it can exhibit a far better metrological performance. Furthermore, the reference profiles for the generation of eLAMI were constructed to take into account the variability of the data over a time horizon longer than 1 hour (total test duration). Of course, increasing the acquisition time would tend to increase the variation intervals of the WT3000 distributions, due to measurand variability.

An interesting aspect, that can be seen in this figure looking at the WT3000 measured values, is the behaviour of the V_{RMS} (Fig. 7.a)). The 3 loads are simultaneously supplied by the same power source, each through its own power line: such a setting can cause a potential voltage drop. As reported in Fig. 7.c), the *Fan Heater* is the load with the highest absorption: consequently, the V_{RMS} at its ends is the lowest (Fig. 7.a)). Conversely, the WT3000 records the highest voltage at the ends of the “*Fan*”, which is the closest to the aggregate’s one, i.e. the power source. Looking at the voltage behavior of eLAMI, the same trend can be observed.

To demonstrate *(ii)*, only the mean value and standard deviation of a few monitored Features for the “*Fan*” load are reported, for sake of brevity, in Tab. 6. In particular, the considered features are: V_{rms} , I_{rms} , P and S .

The standard deviation values of the measurements recorded by the WT3000, as we expected, are much smaller than those related to eLAMI, due to the simulator design parameters, which had the purpose to replicate a typical commercial smart meter less accurate than the adopted reference (WT3000). Nevertheless, from a measurement point of view, the intervals ($\mu-\sigma$, $\mu+\sigma$) belonging to WT3000 and eLAMI are generally overlapped, demonstrating the validity of the generated dataset.

The validity of the algorithms implemented in the simulator for the generation of eLAMI is evident when analyzing the values and behaviors obtained from the features analyzed in Fig. 7 and Tab. 6. Furthermore, the similarity between the behaviors of the “*Aggregate*” highlights the consistency of the process implemented in the simulator.

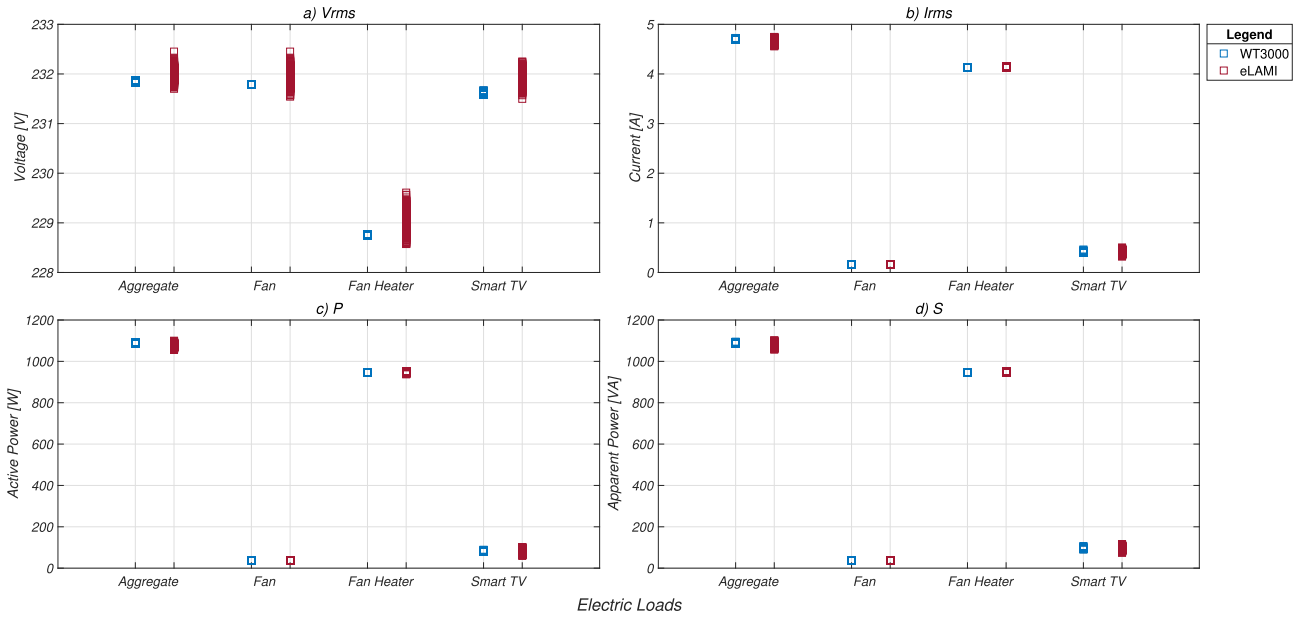


FIGURE 7. Comparison between WT3000 and eLAMI for devices: “Fan”, “Fan Heater”, “Smart TV” and “Aggregate”. Comparison made in terms of features: a) Voltage RMS - b) Current RMS c) Active power d) Apparent power.

TABLE 6. Comparison of mean and standard deviation values obtained from WT3000 and eLAMI for the electrical load “Fan”. Considered features: V_{rms} , I_{rms} , P and S .

Feature		WT3000	eLAMI
V_{rms} [V]	μ	231.68	231.94
	σ	0.13	0.59
I_{rms} [mA]	μ	159.90	161.30
	σ	0.58	1.40
P [W]	μ	36.38	37.12
	σ	0.22	0.37
S [VA]	μ	36.82	37.63
	σ	0.31	0.32

B. DESCRIPTIONS OF THE GENERAL CHARACTERISTICS OF eLAMI

As highlighted above, one of the peculiarities of eLAMI is the variability of the data, both in terms of operating states of the monitored devices and consumption habits. Fig. 8 shows the average monthly active energy consumption profiles of the aggregate load, for the 24 hours of the day (x-axis), for each month. In particular, in terms of active energy consumption (y-axis), the curves show: in green, the average trend, in red the maximum reached for each hourly interval of the month considered, and similarly in blue the minimum. First of all, when analyzing the individual month, we can see the variability of the absorption curve during the 24 hours of the day, consistent with what happens in the residential area. In particular, the curve shows an increase in the early morning hours followed by a rapid decrease until midday when the second absorption peak occurs. A further decrease

follows this in consumption before arriving at the evening hours characterized by the highest energy absorption.

In addition to the variability during the day, by comparing the different months, we can value the seasonality of consumption and thus the variation in electrical-behavioral habits. In particular, consumption is higher during the winter months than in the summer months, which is particularly evident when comparing August and December. This is because in the winter months, according to defined habits, there is greater use of certain high-consumption devices, such as *Boiler*, *Electric Oven* and *Microwave*. Furthermore, it should be noted that during the winter months lamp utilization is higher than in the summer months, which has an impact on the consumption peaks mentioned above.

A further interesting aspect is a non-zero consumption during nighttime hours present in all months, due to the presence of some devices in standby mode, characterized by minimal but not zero power consumption. This ceiling does not show much variability in terms of consumption, as the devices on stand-by during the night are almost always the same, so the 3 curves (average, minimum and maximum) almost overlap, with a consumption of less than 0.5 kWh.

In Fig.9 the total electricity consumption for each month of the simulated year is shown. In this case, the seasonality of consumption and thus the variation of the energy absorbed in the different months is particularly evident. Since there is no cooling system, total consumption is higher in autumn and winter than in spring and summer. In particular, starting the year in January, electricity consumption is high and remains more or less constant until March, and then begins to decrease in April with the arrival of spring. The lowest peak is reached in August and then starts to rise again from September forward. The highest electricity

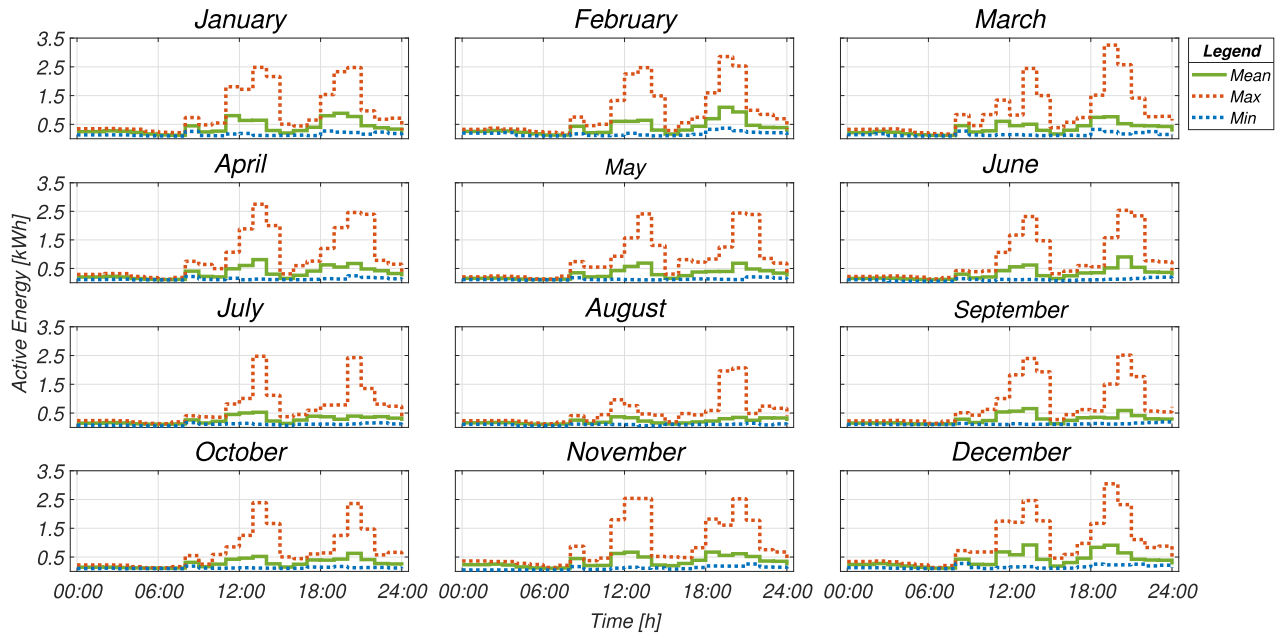


FIGURE 8. Average monthly active energy consumption profiles of the aggregate load.

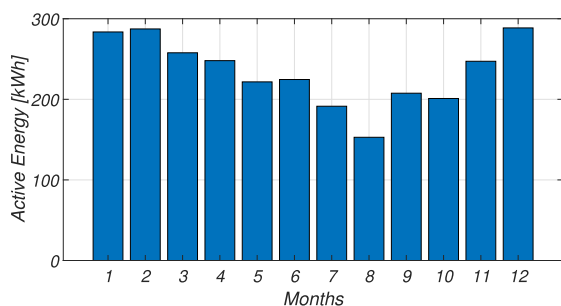


FIGURE 9. Aggregate active energy load curve per months.

consumption is recorded in December, while the lowest is recorded in August in perfect analogy with what is shown in Fig. 8.

In Fig. 10 we report the average daily consumption of 4 electrical loads for each month of eLAMI (in terms of average, maximum and minimum daily consumption), taken as an example, to show the variation in consumption patterns in eLAMI. Specifically, the devices are: *Boiler* (10.a), *Smartphone Charger* (10.b), *Fan* (10.c), *Fan Heater* (10.d). Analyzing *Boiler* consumption (10.a), it can be seen that the average consumption in winter is higher than in summer, where it is used only for hot water and not for heating. In contrast, the *Fan Heater* (10.d) is only used in winter, from October to April, peaking in February. The opposite behavior is obtained by analyzing the *Fan* (10.c), which is only used in the summer months, from June to September, with a peak in August. Unlike the others, the *Smartphone Charger* (10.b) does not show substantial variations in consumption between months. This is because it is used on average every day of the year in the same way.

C. SMART ENERGY APPLICATION EXAMPLES

1) LOAD PROFILING

Machine Learning and Artificial Intelligence techniques, in general, are based on the use of large amounts of input data. However, if these techniques have input data that do not correctly describe the phenomenon to be studied, the output may be far from the desired result. This is why feature selection algorithms are very often used to find the best set of features to build useful and robust models of the phenomena studied [45].

In Fig. 11.a the active power absorbed by the *Desk Lamp* in January eLAMI is reported as a function of the corresponding assumed states. As previously verified, state variability is present. Consequently, the active power P alone is not able to discriminate the 4 different load operating state because states 1 and 2, and similarly 3 and 4, overlap in terms of P . Therefore, in terms of load profiling, other features must be found for the correct identification of operating states. For example, the active power P as a function of the power factor at fundamental is reported in Fig. 11.b. This feature is able to discriminate states 1 and 2 better than P alone, while states 3 and 4 still remain indistinguishable. Conversely, in Fig. 11.c we see the phase of the 6th harmonic voltage group identify states 3 and 4 well but not the first two. Combining the two features identified, power factor at fundamental and phase of the 6th harmonic voltage group, Fig. 11.d shows an optimal situation where 4 operating states of the load (cluster) considered are clearly visible. It is therefore very clear how the greater number of electrical parameters in eLAMI results in a better representation and distinction between the different electrical signatures of the individual load.

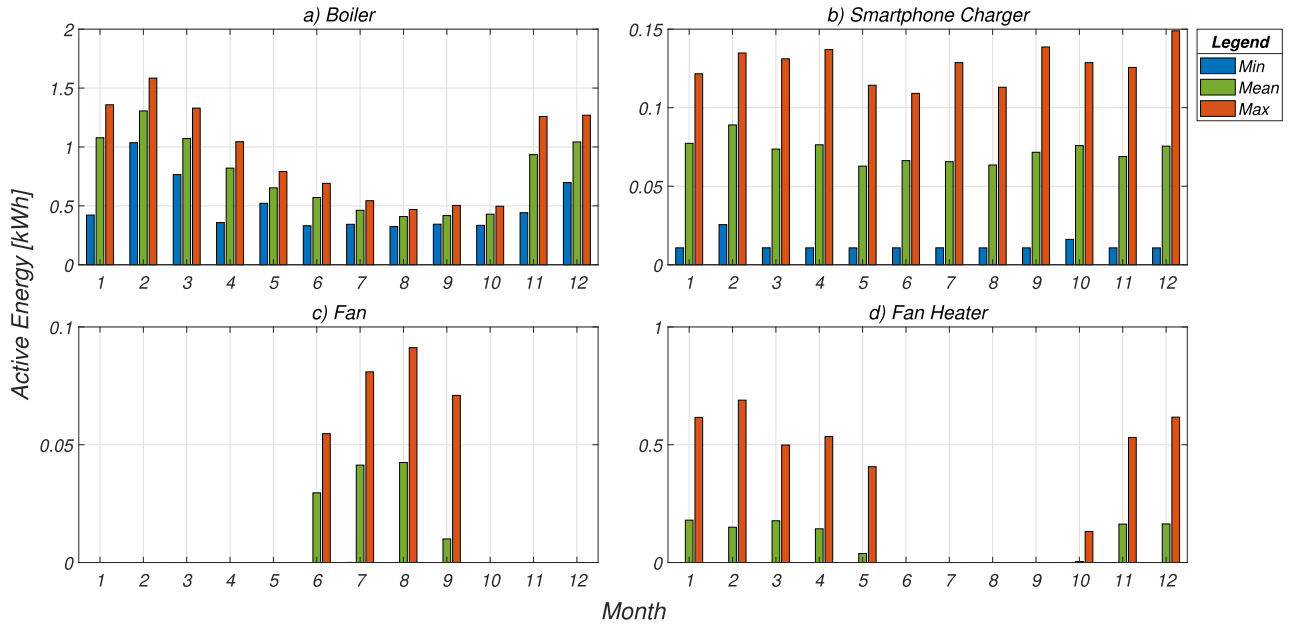


FIGURE 10. Seasonality example of some different electrical loads - a) Boiler - b) Smartphone charger - c) Fan - d) Fan heater.

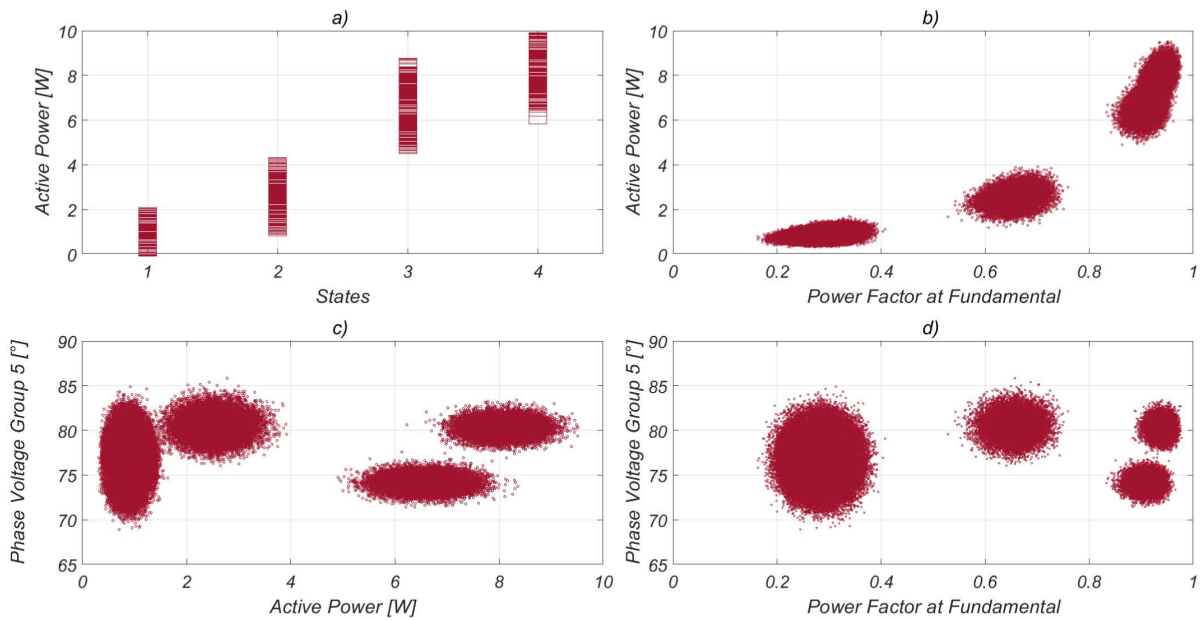


FIGURE 11. Example of load profiling application for desk lamp - a) Power consumption per state b) Active power as a Function of power factor at 50 Hz c) Phase of 5th voltage Harmonic group as a function of active power d) Phase of 5th voltage harmonic group as a function of power factor at 50 Hz.

2) NILM

Below is reported an example of application in the NILM (Non-intrusive Load Monitoring) field. NILM is presented as a time series classification problem where the objective is to detect which appliances are active at a given instant and how much each of them contributes to the total percentage of consumption. Due to their advantages, techniques based on the analysis of steady-state features are the most widely used,

typically referring to active power only [46]. This feature, however, is not always able to distinguish devices that absorb similar power or have similar operating principles.

In Fig. 12.a we report the total active power obtained from an aggregation of 3 loads present in eLAMI and in Fig. 12.b the individual active powers absorbed by the 3 considered loads, are reported. In particular, in Fig. 12.a and Fig. 12.b the active power in time, both for aggregate and some example of

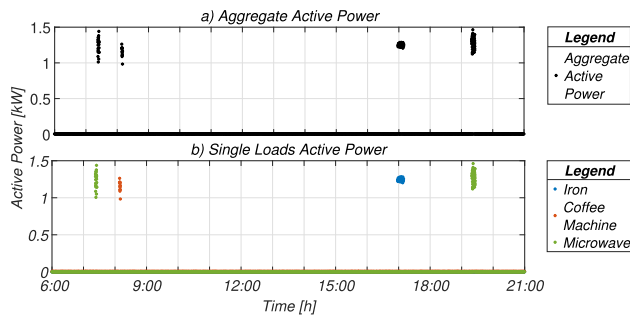


FIGURE 12. Active power in time: a) Aggregate - b) Single loads.

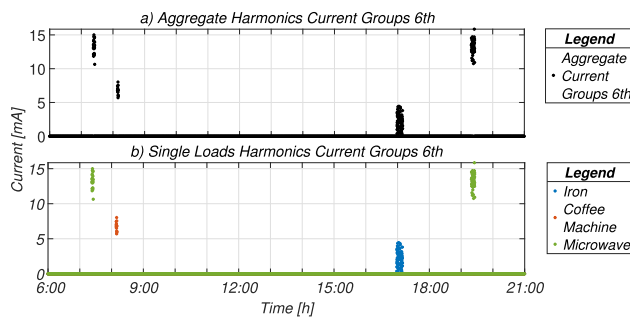


FIGURE 13. Current harmonics of 6th group in time: a) Aggregate - b) Single loads.

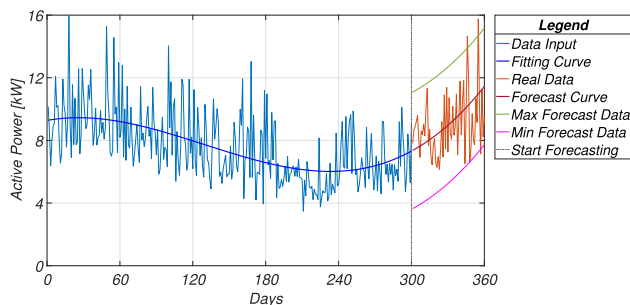


FIGURE 14. Example of forecasting for electric consumption with eLAMI.

single load, are reported. Conversely, Fig. 13.a and Fig. 13.b show the RMS values of the 6th harmonic current group in time, both for the individual loads considered above and for the corresponding aggregate.

It is evident when analyzing the aggregate active power in Fig. 12.a and the single active power in Fig. 13.b, that the 3 loads are indistinguishable due to the problems mentioned above. This represents a critical case for NILM algorithms. Instead, analyzing the current harmonics of 6th group in Fig. 13.a for the aggregate and in Fig. 13.b for every single load, three different levels of absorption are present, allowing for correct identification of the active loads.

3) FORECASTING

In the world of Smart Energy, statistical and forecasting analyses based on time series are often carried out. In particular,

one of the goals of an Energy Management System is to create mathematical models that can simulate trends in electricity consumption as a function of various factors. Furthermore, through statistical analysis, it is possible to define statistical and/or performance indices of the analyzed system. Some of the benefits of modeling analyzed electricity consumption are: *i*) construction of past seasonal trends and consequently of future ones, *ii*) definition of energy efficiency and optimization plans, *iii*) forecast balancing of electricity networks and performance verification.

To this end, eLAMI also offers the possibility of testing forecasting and modeling algorithms for electrical systems. For example, in Fig. 14 based on the knowledge of the first 10 months of the year (300 days) relating to the power consumption trend of the total aggregate, a consumption forecast was made for the last two months of the year. The mathematical model derived is based on a fitting of the input data. Using the mathematical model, a band ($2 * DevStd$ width) was derived within which the consumption for the 60 forecast days is estimated. As can be seen from this figure, the real data falls within the obtained forecast band with an error of approximately 6.6 %, i.e. 56 days out of 60 estimated correctly.

Furthermore, in diagnostic and predictive terms, thanks to the analysis of time series of data, it is possible to estimate the operating ranges of electrical loads in order to assess and/or predict any decay in their performance and any drift towards fault states.

V. CONCLUSION

A novel approach to provide a simulated electrical energy dataset, from the reference signal acquisition to the data validation has been reported in this paper. To enforce motivations leading to its building and to prove its suitability for Smart Energy applications, a final section regarding examples of Smart Energy profiling and management is also discussed. The output data, composed of more than 400 electrical parameters and reporting one-year period energy profile simulation of a residential building, are made available for download to enhance research in Smart Energy sector with novel, detailed, validated and wide-applicable data. The acquisition set-up has been chosen according to typical metering capabilities of currently adopted home smart meters. Voltage and Current signals, produced for 36 home appliances, have been processed both in time and frequency domain, to provide a comprehensive set of electrical parameters composing the electrical signature of each considered appliance. To be as close as possible to the real scenario, stochastic models have been also implemented to obtain consumption habits to manage state transitions for each load. In its current status, only nominal operating conditions have been considered, i.e. no failures have been hypothesized during the simulation interval. This is, to the authors' opinion, still reasonable given the very low failure rate of the considered apparatuses in the tested period. The produced data are anyway suitable for most Smart

Metering Applications. A second release is intended to be developed, where common failures will be implemented and its aim would be voted to fault location research efforts: it could be seen as an appendix to the current dataset, which results as much complete as possible in terms of detailed representation of typical operating states of considered appliances in a home environment.

REFERENCES

- [1] COP26. (2022). *COP26: The Glasgow Climate Pact*. Accessed: Mar. 10, 2022. [Online]. Available: <https://ukcop26.org/wp-content/uploads/2021/11/COP26-Presidency-Outcomes-The-Climate-Pact.pdf>
- [2] CE. (2022). *Directive 2006/32/EC of the European Parliament and of the Council*. Accessed: Mar. 14, 2022. [Online]. Available: <https://eur-lex.europa.eu/legal-content/IT/TXT/PDF/?uri=CELEX:32006L0032&from=LV>
- [3] CE. (2022). *Directive 2009/72/EC of the European Parliament and of the Council*. Accessed: Mar. 14, 2022. [Online]. Available: <https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32009L0072&from=EN>
- [4] CE. (2022). *Directive 2012/27/CE of the European Parliament and of the Council*. Accessed: Mar. 14, 2022. [Online]. Available: <https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32012L0027&from=IT>
- [5] K. A. Agyeman, S. Han, and S. Han, "Real-time recognition non-intrusive electrical appliance monitoring algorithm for a residential building energy management system," *Energies*, vol. 8, no. 9, pp. 9029–9048, 2015.
- [6] W. Kong, Z. Y. Dong, J. Ma, D. J. Hill, J. Zhao, and F. Luo, "An extensible approach for non-intrusive load disaggregation with smart meter data," *IEEE Trans. Smart Grid*, vol. 9, no. 4, pp. 3362–3372, Jul. 2018.
- [7] Y. F. Wong, Y. A. Şekerçioğlu, T. Drummond, and V. S. Wong, "Recent approaches to non-intrusive load monitoring techniques in residential settings," in *Proc. IEEE Comput. Intell. Appl. Smart Grid (CIASG)*, Apr. 2013, pp. 73–79.
- [8] V. Stankovic, J. Liao, and L. Stankovic, "A graph-based signal processing approach for low-rate energy disaggregation," in *Proc. IEEE Symp. Comput. Intell. Eng. Solutions (CIES)*, Dec. 2014, pp. 81–87.
- [9] B. Zhao, L. Stankovic, and V. Stankovic, "On a training-less solution for non-intrusive appliance load monitoring using graph signal processing," *IEEE Access*, vol. 4, pp. 1784–1799, 2016.
- [10] G. Berrettoni, C. Bourelly, D. Capriglione, L. Ferrigno, and G. Miele, "Preliminary sensitivity analysis of combinatorial optimization (CO) for NILM applications: Effect of the meter accuracy," in *Proc. IEEE 6th Int. Forum Res. Technol. Soc. Ind. (RTSI)*, Sep. 2021, pp. 486–490.
- [11] H.-H. Chang, "Non-intrusive demand monitoring and load identification for energy management systems based on transient feature analyses," *Energies*, vol. 5, no. 11, pp. 4569–4589, 2012.
- [12] Q. Wu and F. Wang, "Concatenate convolutional neural networks for non-intrusive load monitoring across complex background," *Energies*, vol. 12, no. 8, p. 1572, 2019.
- [13] J. Kim, T.-T.-H. Le, and H. Kim, "Nonintrusive load monitoring based on advanced deep learning and novel signature," *Comput. Intell. Neurosci.*, vol. 2017, Oct. 2017, Art. no. 4216281.
- [14] S. Singh and A. Majumdar, "Deep sparse coding for non-intrusive load monitoring," *IEEE Trans. Smart Grid*, vol. 9, no. 5, pp. 4669–4678, Sep. 2017.
- [15] L. Mauch and B. Yang, "A new approach for supervised power disaggregation by using a deep recurrent LSTM network," in *Proc. IEEE Global Conf. Signal Inf. Process. (GlobalSIP)*, Dec. 2015, pp. 63–67.
- [16] V. M. Salerno and G. Rabbeni, "An extreme learning machine approach to effective energy disaggregation," *Electronics*, vol. 7, no. 10, p. 235, 2018.
- [17] S. Giri, M. Bergés, and A. Rowe, "Towards automated appliance recognition using an EMF sensor in NILM platforms," *Adv. Eng. Informat.*, vol. 27, no. 4, pp. 477–485, 2013.
- [18] A. Ruano, A. Hernandez, J. Ureña, M. Ruano, and J. Garcia, "NILM techniques for intelligent home energy management and ambient assisted living: A review," *Energies*, vol. 12, no. 11, p. 2203, Jun. 2019.
- [19] Y. Wang, Q. Chen, C. Kang, M. Zhang, K. Wang, and Y. Zhao, "Load profiling and its application to demand response: A review," *Tsinghua Sci. Technol.*, vol. 20, no. 2, pp. 117–129, Apr. 2015.
- [20] J. Yan, Y. Meng, L. Lu, and L. Li, "Industrial big data in an industry 4.0 environment: Challenges, schemes, and applications for predictive maintenance," *IEEE Access*, vol. 5, pp. 23484–23491, 2017.
- [21] S. Makonin, F. Popowich, L. Bartram, B. Gill, and I. V. Bajić, "AMPds: A public dataset for load disaggregation and eco-feedback research," in *Proc. IEEE Electr. Power Energy Conf.*, Aug. 2013, pp. 1–6.
- [22] S. Makonin, B. Ellert, I. V. Bajić, and F. Popowich, "Electricity, water, and natural gas consumption of a residential house in Canada from 2012 to 2014," *Sci. Data*, vol. 3, no. 1, pp. 1–12, 2016.
- [23] A. Filip, "BLUED: A fully labeled public dataset for event-based nonintrusive load monitoring research," in *Proc. 2nd Workshop Data Mining Appl. Sustainability*, 2011, pp. 1–5.
- [24] O. Parson, G. Fisher, A. Hersey, N. Batra, J. Kelly, A. Singh, W. Knottenbelt, and A. Rogers, "Dataport and NILMTK: A building data set designed for non-intrusive load monitoring," in *Proc. IEEE Global Conf. Signal Inf. Process. (GlobalSIP)*, Dec. 2015, pp. 210–214.
- [25] A. S. U. Nambi, A. R. Lua, and V. R. Prasad, "LocED: Location-aware energy disaggregation framework," in *Proc. 2nd ACM Int. Conf. Embedded Syst. Energy-Efficient Built Environ.*, 2015, pp. 45–54.
- [26] C. Beckel, W. Kleiminger, R. Cicchetti, T. Staake, and S. Santini, "The ECO data set and the performance of non-intrusive load monitoring algorithms," in *Proc. 1st ACM Conf. Embedded Syst. Energy-Efficient Building*, 2014, pp. 80–89.
- [27] C. Shin, E. Lee, J. Han, J. Yim, W. Rhee, and H. Lee, "The enertalk dataset, 15 Hz electricity consumption data from 22 houses in Korea," *Sci. Data*, vol. 6, no. 1, pp. 1–13, 2019.
- [28] N. Batra, M. Gulati, A. Singh, and M. B. Srivastava, "It's different: Insights into home energy consumption in India," in *Proc. 5th ACM Workshop Embedded Syst. Energy-Efficient Building*, 2013, pp. 1–8.
- [29] J. Z. Kolter and M. J. Johnson, "REDD: A public data set for energy disaggregation research," in *Proc. Workshop Data Mining Appl. Sustainability (SIGKDD)*, vol. 25. San Diego, CA, USA, 2011, pp. 59–62.
- [30] D. Murray, L. Stankovic, and V. Stankovic, "An electrical load measurements dataset of United Kingdom households from a two-year longitudinal study," *Sci. Data*, vol. 4, no. 1, pp. 1–12, 2017.
- [31] J. Kelly and W. Knottenbelt, "The UK-DALE dataset, domestic appliance-level electricity demand and whole-house demand from five UK homes," *Sci. Data*, vol. 2, no. 1, pp. 1–14, 2015.
- [32] N. Buneeva and A. Reinhardt, "AMBAL: Realistic load signature generation for load disaggregation performance evaluation," in *Proc. IEEE Int. Conf. Smart Grid Commun. (SmartGridComm)*, Oct. 2017, pp. 443–448.
- [33] S. Henriët, U. Şimşekli, B. Fuentes, and G. Richard, "A generative model for non-intrusive load monitoring in commercial buildings," *Energy Buildings*, vol. 177, pp. 268–278, Oct. 2018.
- [34] C. Klemenjak, C. Kovatsch, M. Herold, and W. Elmenreich, "A synthetic energy dataset for non-intrusive load monitoring in households," *Sci. Data*, vol. 7, no. 1, pp. 1–17, 2020.
- [35] D. Chen, D. Irwin, and P. Shenoy, "SmartSim: A device-accurate smart home simulator for energy analytics," in *Proc. IEEE Int. Conf. Smart Grid Commun. (SmartGridComm)*, Nov. 2016, pp. 686–692.
- [36] eLAMI. (2022). *An Innovative Simulated Dataset of Electrical Loads for Advanced Smart Energy Applications*. Accessed: Jul. 11, 2022. [Online]. Available: <https://ieee-dataport.org/documents/elami>
- [37] Pacific Smart Source. (2022). *Pacific 360 AMX*. Accessed: Feb. 2, 2022. [Online]. Available: https://pacificpower.com/Resources/Documents/360AMX_Datasheet.pdf
- [38] Tektronix. (2022). *TCP202 DC Coupled Current Probe*. Accessed: Feb. 2, 2022. [Online]. Available: https://download.tek.com/datasheet/TCP202_Current_Probe_Datasheet_51W-10736-7.pdf
- [39] Tektronix. (2022). *Probe Power Supply 1103*. Accessed: Feb. 2, 2022. [Online]. Available: https://download.tek.com/datasheet/1103_Probe_Power_Supply_Datasheet_60W-14843-1_2.pdf
- [40] Tektronix. (2022). *High Voltage Differential Probes*. Accessed: Feb. 2, 2022. [Online]. Available: <https://download.tek.com/manual/P5200A-High-Voltage-Differential-Probe-Instruction-Manual-077053602.pdf>
- [41] TiePie. (2022). *Handyscope HS5*. Accessed: Feb. 2, 2022. [Online]. Available: <https://www.tiepie.com/en/usb-oscilloscope/handyscope-hs5>
- [42] *IEEE Standard Definitions for the Measurement of Electric Power Quantities Under Sinusoidal, Nonsinusoidal, Balanced, or Unbalanced Conditions*. IEEE Standard 1459-2010, 2010, pp. 1–50.
- [43] (2022). *IEC 61000-4-7*. Accessed: Feb. 7, 2022. [Online]. Available: <https://webstore.iec.ch/publication/4228>

- [44] Yokogawwa. (2022). *WT3000 Yokogawa*. Accessed: Apr. 2, 2022. [Online]. Available: <https://tmi.yokogawa.com/solutions/discontinued/wt3000-precision-power-analyzer/>
- [45] J. Tang, S. Alelyani, and H. Liu, *Feature Selection for Classification: A Review*. CRC Press, Jan. 2014, pp. 37–64.
- [46] G. W. Hart, “Nonintrusive appliance load monitoring,” *Proc. IEEE*, vol. 80, no. 12, pp. 1870–1891, Dec. 1992.



LUCA TARI (Graduate Student Member, IEEE) was born in Cassino, Italy, in 1997. He received the M.S. degree (*cum laude*) in electrical engineering from the University of Cassino and Southern Lazio, Italy, in October 2021. He is currently pursuing the Ph.D. degree with the Department of Electrical Engineering, University of Cassino and Southern Lazio. His current research interests concern electrical and electronic measurements in the field of Smart Energy. In particular, the design of innovative Smart Meters and the development of methods and techniques for monitoring, analysis, energy efficiency, fault detection and predictive diagnostics.



GIUSEPPE BERRETTONI (Graduate Student Member, IEEE) received the master’s degree in electrical engineering from the University of Cassino and Southern Lazio, Cassino, Italy, in 2019, where he is currently pursuing the Ph.D. degree in methods, models and technologies for engineering, where his research activity is focused on measurement for smart energy application, in particular on analysis of installation non idealities for current measurement in MV/LV cabin with Rogowski coils, smart metering for NILM, load profiling, and predictive maintenance applications in industry 4.0. He has got a research grant from the University of Cassino, from May 2020 to November 2020.



CARMINE BOURELLY (Graduate Student Member, IEEE) received the master’s degree in electrical engineering from the University of Cassino and Southern Lazio, Cassino, Italy, in 2017, where he is currently pursuing the Ph.D. degree, where his main interests are focused on battery state of health estimation. His research interests include detection of pollutants in wastewater, systems for measuring radon concentration in air and evaluation of meter accuracy in NILM applications. He worked as an Engineer in research and development area at Italian Company Sensichips s.r.l.



GIANNI CERRO (Member, IEEE) is currently a Research Fellow with the Department of Medicine and Health Sciences “Vincenzo Tiberio”, University of Molise, Italy. His research interests include magnetic localization systems for biomedical and industrial applications, cognitive radio systems for new generation communication technologies, measurements in telecommunication networks, sensor networks for environmental monitoring, and measurement characterization of medical devices, such as brain–computer interfaces.



DOMENICO CAPRIGLIONE (Senior Member, IEEE) is currently an Associate Professor of electrical and electronic measurements with the University of Cassino and Southern Lazio, Italy. His current research interests include measurements on RF and telecommunication systems, DSP-based measurement systems, network measurements, and measurement of electromagnetic compatibility. Since 2016, he has served as the Chair for the IEEE I&M Technical Committee TC-37-Measurements and Networking.



LUIGI FERRIGNO (Senior Member, IEEE) has been a Full Professor of electric and electronic measurement and the Scientific Manager of the Industrial Measurements Laboratory, University of Cassino, Cassino, Italy, since 2004. In 2008, he was a Founding Member of the university spin-off, Spring Off (University of Salerno), Fisciano, Italy. He is an NDE4.0 Ambassador for the Italian Association of Non-Destructive Evaluation and Test (AiPnD) in the EFNDT WG10. He has coordinated and participated in several national and international research projects. His current research interests include the NDT4.0, novel learning sensors and measurement systems for smart city, the Internet of Things (IoT), automotive, smart energy, and environment.

• • •